*THEIL (General Entropy) INDEXES

A. Concept: Theil-T index assess how much a given income distribution (each person receive y_i of total income) is away of a perfect uniform distribution (each person receive 1/n of total income), or the redundancy degree in relation to the latter, weighting each observation by its share in total income.

$$T = \ln n - H(x) = \sum_i y_i \ln \frac{y_i}{1/n}$$

$0 \le T \le \ln n$, that is, we have $T = 0$ in the case of a perfect egalitarian distribution and $T = \ln n$ in the case of maximum inequality. Theil-T index assess how much a given income distribution (each person receive y_i of total income) is away of a perfect uniform distribution (each person receive 1/n of total income), or the redundancy degree in relation to the latter, weighting each observation by its share in total income. If in ln in *nits (natural logs units),*

The second Theil measure of inequality is Theil-L index, defined by the following formula:

$$L = \sum_{i=1}^{n} \frac{1}{n} \log \frac{1/n}{y_i} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{1/n} \ (-1)$$

It inverts the redundancy comparison and weights. While in Theil T the inequality factors of weighting within the groups are the share of income, in Theil L the inequality factors of weighting within the groups are their respective population.

or, alternatively, by $T = \sum_{i=1}^{n} \frac{x_i}{N\mu} \log \frac{x_i}{\mu}$ that is comparing means instead of shares

B. Intra and Inter Groups Decomposition of Theil T (Theil L allows a similar formula)

$T = T_e + \sum_{h=1}^{K} Y_h T_h$ Where, $T_e = \sum_{h=1}^{k} Y_h \log \frac{Y_h}{\pi_h}$ is the Theil T between groups and $T_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log n_h \frac{y_{hi}}{Y_h}$ is the Theil intra groups. Therefore $\sum_{h=1}^{K} Y_h T_h$ is the weighted average of intra-groups Theil Ts. **Te / T is the Contribution of a certain characteristic to inequality (say how much schooling (or gender) explains <u>exactly</u> total inequality?). Alternative to mincerian regressions based decompositions.**

Theil Between groups ← **Theil Total** → **Theil Within groups**

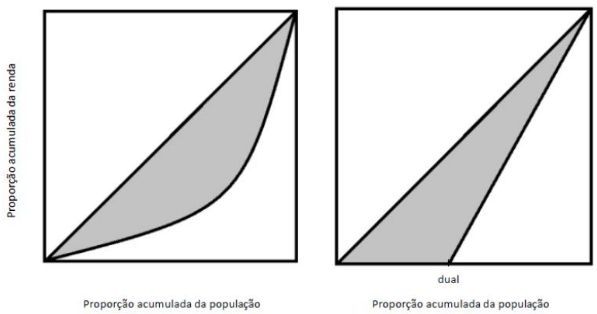
Other application: Does per capita Household Income underestimates true inequality?

Theil Between Families (per capita) ← **Theil Total** → **Theil Within Families**

***Applying Decomposition to Inequality & Temporal Variability (Mobility, Risk or measurement error)**

Dispersion of Mean Income (earned over time) between people ← **Theil Total** → **Mean variability of income in relation to the mean across time for each person**

C. Dual: $U_2 = \phi + (1-\phi)U_1$ allows to compare different inequality measures in the same 0 to 1 scale **The Dual of the Gini Index is the Gini Index G* = G (1-%) + %, % are** new 0s a way to proceed with maximum inequality (G=1) so is adding top incomes. One can use this formula for introducing both ends of income distribution. As the dual of any inequality measure since its dual transformation measures in the Gini scale. Applying this formula $U_2 = \phi + (1-\phi)U_1$ to the to the Theil –T we get $T2 = T1 - \ln(1-\phi)$. A fully decomposable overall measure of social welfare inspired on Sen (1973) is $SW = mean.(1-U_{T1})$. Since the Theil L does not admit null values, it also does not admit a Dual measure. **The comparison of the Lorenz Diagrams below captures the idea of the Dual:**



D. General Entropy S- measure nests Theil T and Theil L among other indexes, as special cases.

According to the general formula of a inequality measure

$$S = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\mu}\right)^{1-\varepsilon}\right]$$

OBS: **ε=0 Theil T; ε=1 Theil L; ε= -1 Coefficient of Variation**

- $\varepsilon = -1$

We have that

$S = \frac{1}{2}\left[\frac{1}{N}\sum_{i=1}^{N}(\frac{x_i}{\mu})^2 - 1\right]$

$2S = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2}{\mu^2} = CV^2 \implies S = \frac{CV^2}{2}$

That is, when $\varepsilon = -1$ the measure is half of the square of the coefficient of variation.

- $\varepsilon = 0$

In this case, the value of S is not defined. We have to use L'Hôpital to calculate the limit when $\varepsilon \to 0$. Note that

$limS = \frac{1}{1-2\varepsilon}\frac{1}{n}\sum(\frac{x_i}{\mu})^{1-\varepsilon}ln(\frac{x_i}{\mu})$

When $\varepsilon \to 0$, we have that

$limS = \frac{1}{n}\sum(\frac{x_i}{\mu})ln(\frac{x_i}{\mu}) = T$

That is, when $\varepsilon \to 0$, the measure approaches the Theil-T.

- $\varepsilon = 1$

Once again we will have to use L'Hôpital. When $\varepsilon \to 1$, we have that

$limS = -\frac{1}{n}\sum ln(\frac{x_i}{\mu}) = L$

That is, when $\varepsilon \to 1$, the measure approaches the Theil-L.

C. Detailing the Dual *(Hoffman 1991 book pages 42-44 and 107-110)

Dual General Definition:

Be x a random variable with mean μ and distribution with certain value of inequality as **M**. We called *dual* a distribution with the following characteristics:

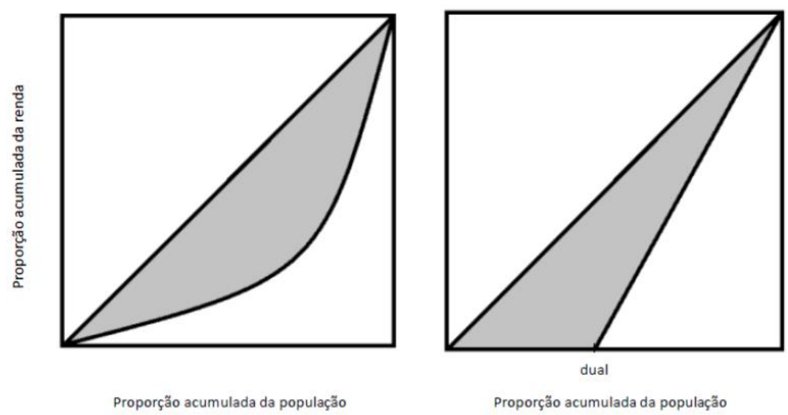
a. $x = 0$ with probability U_t and $x = \mu / (1- U_t)$ with probability $1 - U_t$. That is, maintain the original mean for any $U_{t..}$ The inequality measure value is also equal to M, once we adjusted U_t value.

Dual maintain the mean and inequality for the value U_t.. Dual allows different comparisons of inequality measures.

Main advantages:

a) identical scales and vary in the interval 0 to 1, (same as Gini's), dimensionless
b) allows to study the sensitivity of the measure of inequality
c) allows equivalence between measures.

The comparison of the Lorenz Diagrams below captures the idea of the Dual:



Cummulative Share in the population **Cummulative Share in the population**

Given the relationship of the Lorenz curve with the geometrical interpretation of the Gini index: **The Dual of the Gini Index is the Gini Index G* = G (1-%) + %, % are** new 0s a way to proceed with maximum inequality (G=1) so is adding top incomes. One can use this formula for introducing both ends of income inequality. As the dual of any inequality measure since its dual transformation measures in the Gini scale.

$U_2 = \phi + (1-\phi)U_1$

Deduction of the Dual from the Theil-T Index

In terms of the fraction of the total income of the population received by each person, in the dual distribution we have

$y_i = 0$, for nU_T people, and

$y_i = \dfrac{1}{n(1-U_T)}$, for $n(1-U_T)$ people

Thus, according to the formulas given above, we have:

$$T = \sum_{i=1}^{n} y_i \log ny_i = nU_T \left[0 \log n0\right] + n(1-U_T)\left[\frac{1}{n(1-U_T)} \log n \frac{1}{n(1-U_T)}\right] = \log \frac{1}{(1-U_T)}$$

Raising to exponential, we obtain:

$$e^T = \frac{1}{(1-U_T)} \Rightarrow 1-U_T = e^{-T} \Rightarrow U_T = 1-e^{-T}$$

$$0 \le T \le \log n$$
$$1 \le e^{T} \le n$$
$$1 \ge e^{-T} \ge \frac{1}{n}$$
$$-1 \le -e^{-T} \le -\frac{1}{n}$$
$$0 \le 1-e^{-T} \le 1-\frac{1}{n}$$
$$0 \le U_T \le 1-\frac{1}{n}$$

A dual distribution follows the equation below:

$$U_2 = \phi + (1-\phi)U_1$$

Where U_1 is the dual of the initial distribution and U_2 is the dual after adding null values that are a proportion $\phi = \frac{m}{n+m}$ of the new total elements. Thus, for the Theil-T we have:

$$U_{T2} = \phi + (1-\phi)U_{T1}$$

What bring us to:

$$1-e^{-T2} = \phi + (1-\phi)(1-e^{-T1})$$
$$1-e^{-T2} = \phi + (1-\phi) - (1-\phi)e^{-T1}$$
$$e^{-T2} = (1-\phi)e^{-T1}$$
$$-T2 = \ln(1-\phi) - T1$$

$$T2 = T1 - \ln(1-\phi)$$

Where $T1$ and $T2$ are values, in *nits (natural logs units)*, of the Theil-T index for the initial distribution and after the adding of the *m* set of null values, respectively.

B. **Deduction of Intra and Inter Groups Decomposition of the Theil T

Suppose I have a population with N samples, divided in K groups:

$N = \sum_{h=1}^{K} n_h$, which n_h is the nº of people in the h-th group. The proportion of the population correspondent to the h-th group would be: $\pi_h = \frac{n_h}{N}$.

Suppose that x_{hi} is the i-th individual income of the h-th group. Thus, total income share of this individual would be: $y_{hi} = \frac{x_{hi}}{N\mu}$, note that the denominator is the population total income, with μ as the mean income. So, the share of the total income retained by the h-th group is:

$Y_h = \sum_{i=1}^{n_h} y_{hi}$, that is, adding the share of total income retained by the individuals within group h.

We have Theil-T Index:

$T = \sum_{i=1}^{N} y_i \log N y_i = \sum_{h=1}^{k}\sum_{i=1}^{n_h} y_{hi} \log N y_{hi}$, Firstly, I'm only first the individuals within the group, and then adding the others until complete all the population.
Adding and subtracting:

(*) $\sum_{h=1}^{k} Y_h \log \frac{NY_h}{n_h} = \sum_{h=1}^{k}\sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$ (from left to right, I opened Y_h which is out of the log, as defined above ($Y_h = \sum_{i=1}^{n_h} y_{hi}$). Thereby, the equation turn to:

$T = \sum_{h=1}^{k} Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^{k}\sum_{i=1}^{n_h} \frac{Y_h}{Y_h} y_{hi} \log N y_{hi} - \sum_{h=1}^{k}\sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$, which I added and subtracted (*) and divided and multiplied for Yh. Continuing:

$$T = \sum_{h=1}^{k} Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^{k} Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log N y_{hi} - \sum_{h=1}^{k}\sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$$

$$T = \sum_{h=1}^{k} Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^{k} Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log N y_{hi} - y_{hi} \log \frac{NY_h}{n_h} \right]$$

$$T = \sum_{h=1}^{k} Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^{k} Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{N y_{hi}}{\frac{NY_h}{n_h}} \right]$$

$$T = \sum_{h=1}^{k} Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^{k} Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{n_h y_{hi}}{Y_h} \right]$$

$$T = T_e + \sum_{h=1}^{K} Y_h T_h$$

$T = T_e + \sum_{h=1}^{K} Y_h T_h$ Where, $T_e = \sum_{h=1}^{k} Y_h \log \frac{Y_h}{\pi_h}$ is the Theil between groups and $T_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log n_h \frac{y_{hi}}{Y_h}$ is the Theil intra groups. Therefore $\sum_{h=1}^{K} Y_h T_h$ is the weighted average of intra-groups Theil Ts.

Te / T is the Contribution of a certain characteristic to inequality.

Theil Total	**=**	**Theil Between groups**	**+**	**Theil Within groups**

Similarly, Theil-L can be decomposed as between groups (Le) and within groups terms: $L = Le + \sum_{h=1}^{k} \pi h \, Lh$

Where $Le = \sum_{h=1}^{k} \pi h \log(\pi h \,/\, Yh)$ & $Lh = \frac{1}{nh} \sum_{i=1}^{k} \pi h \log(Yh \,/\, (nh\, yhi))$

****GROSS RATES OF CONTRIBUTION THEIL-T**
Universe : Per Capita - All Income Sources

	GROSS							
	1976	**1985**	**1990**	**1993**	**1997**	**2002**	**2003**	**2004**
Groups:								
Gender	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Race	---	---	11.2%	10.8%	12.1%	10.7%	11.6%	10.2%
Age	0.2%	0.1%	0.2%	0.4%	0.9%	1.7%	2.0%	1.8%
Schooling	***36.6%***	***42.4%***	**40.3%**	***36.8%***	***41.3%***	**38.2%**	**36.7%**	**35.2%**
Working Class	12.0%	15.1%	13.4%	11.9%	14.2%	13.2%	14.7%	13.9%
Sector of Activity	13.7%	11.3%	10.3%	7.8%	10.2%	---	---	---
Population Density	17.6%	13.6%	13.5%	9.1%	11.1%	8.2%	6.7%	6.4%
Region	10.2%	8.4%	8.0%	6.9%	8.3%	7.2%	7.8%	7.0%

Source: PNAD

* Theil-T Decomposition and Concepts: Income and Units of Analysis

The Theil-T is the central measurement used here, considering its exact decomposition property. We will work with five pairs of population-income concepts using PNAD:

Income Concept	Population Concept			
	Occupied	Economically Actives	Active Age	Total
Labor * NH				
Labor				
Individual - All sources				
Per capita - All sources				

*NH = Normalized for Working Hours
As the central reference value, we will use Theil-T based on the all sources of per capita income.

RATES OF CONTRIBUTION THEIL-T - 1997

GROSS RATES

Population Concept	Occupied	Occupied	Economically A	Active Age	Total - Per Capita
Income Concept	**Labor NH1**	**Labor**	**All Sources**	**All Sources**	**All Sources**
Groups:					
Gender	0,6%	2,7%	2,7%	3,3%	0,0%
Race	8,3%	9,4%	9,4%	8,5%	12,1%
Age	6,6%	7,8%	8,2%	7,3%	0,9%
Schooling	35,0%	34,6%	34,7%	36,0%	41,3%
Working Class	16,8%	21,0%	21,4%	19,8%	14,2%
Sector	5,9%	5,1%	5,6%	6,0%	10,2%
Population Density	6,9%	7,5%	7,8%	7,5%	11,1%
Region	4,0%	5,4%	5,4%	4,9%	8,3%

MARGINAL RATES

Population Concept	Occupied	Occupied	Economically A	Active Age	Total - Per Capita
Income Concept	**Labor NH1**	**Labor**	**All Sources**	**All Sources**	**All Sources**
Groups:					
Age	3,9%	4,7%	5,9%	5,7%	2,8%
Schooling	26,6%	25,7%	626,4%	28,0%	34,9%
Working Class	5,6%	8,7%	8,7%	8,5%	5,3%

1/ Normalized by Hours

*Income Inequality and Income Mobility - ANTHONY SHORROCKS

The usual indices of inequality are derived from observations on income, wealth etc. corresponding to a particular point or period of time, It has been frequently argued that inequality values by themselves do not accurately reflect the differences between individuals, since the true situation depends to a large extent on how the relative positions of individuals vary over time. Thus, it has been argued, “static” measures of inequality should be supplemented by “dynamic” measures of changes through time, which we shall call measures of mobility. Studies which have proposed ways of quantifying these dynamic changes broadly fall into two categories: those which use elementary statistics, such as the correlation coefficient; and those which make more sophisticated suggestions based on transition matrices and other simple stochastic specifications of dynamic processes. Shorrocks [9] provides a number of references and discusses some of the issues involved in deriving an index of mobility from transition matrices. Particular consideration is given to the interval of time between observations, since a relationship is expected between the amount of observed movement and the length of time over which movement can take place; in a short space of time there is little opportunity for movement, even if the society is inherently very mobile. These earlier attempts to define an index of mobility are mainly concerned with stock variables, interpreted in a wide sense to include social status and occupation as well as wealth and the assets of firms. Once attention is turned to flow variables, such as income, it becomes apparent that there is another important consideration. Observed variations in income depend not only on the interval between observations, but also on the length of the accounting period chosen for incomes. Data availability and custom dictate that the period selected is normally one year, although shorter intervals, a week or a month, are occasionally used. If the accounting period were extended from, say, one month to one year, variations in monthly incomes (previously classified as dynamic changes) become subsumed within the annual income figure. Some of the dynamic changes are therefore incorporated in the static inequality value, and the distinction between the static and dynamic aspects becomes very blurred. Similarly, as we pass from annual to lifetime income inequality, intra-lifetime income mobility is lost in the process of aggregation. However, the effects of income variations over time do not disappear altogether: they are reflected in the changes recorded in the inequality value. Those occupying the highest and lowest positions in the income hierarchy rarely remain there forever. So the aggregation of incomes over time tends to improve the relative position of those temporarily found at the bottom of the distribution, and the situation of those at the top tends to deteriorate. For this reason it is commonly supposed that inequality falls as the accounting period is lengthened. Empirical confirmation of this relationship requires longitudinal income data samples, of which very few exist. However, the little evidence available agrees with expectati0ns.l For example, Soltow [l0] traced the annual incomes of a sample of Norwegians over the period 1928-l960. The Gini coefficient for the 33 years combined was 0.134 compared to an average value of 0.183 for the separate years. Using US data, Kohen et al. [3] found that the Gini coefficient for family income and earnings of young men (aged 16-24) fell by 4.7-7.4 “/„ when cumulated over two years, and by 9.2-10.8 % when cumulated over three. For middle-aged men (45- 59 years old), aggregating incomes over two years caused the Gini to decline by about 4 %.” There are reasonable grounds, therefore, for supposing that the existence of mobility causes inequality to decline as the accounting interval grows. Furthermore, intuition suggests that the extent to which inequality declines will be directly related to the frequency and magnitude of relative income variations. If the income structure exhibits little mobility, relative incomes will be left more or less unaltered over time and there will be no pronounced egalitarian trend as the measurement period increases. In contrast, inequality may be expected to decrease significantly in a very (income) mobile society. The main purpose is to exploit this relationship between mobility and inequality, to derive an index of mobility for flow variables. In essence, mobility is measured by the extent to which the income distribution is equalized as the accounting period is extended. Defining Mobility as the complement of rigidity, as much as we define equality as the complement of inequality. For inequality measures with the desirable properties.

Rigidity Index = Income Inequality Index for Longer Period/ Mean Inequality Index for Shorter Periods

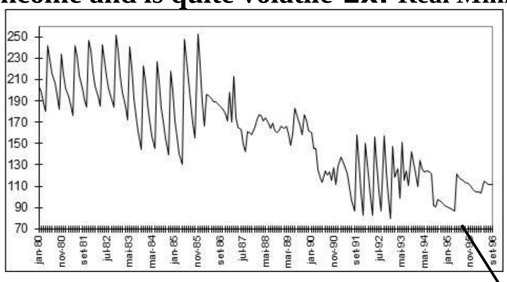
*Applying Decomposition to Temporal Variability (Mobility or Risk)

Dispersion of Mean Income (earned over time) between people

Theil

Mean variability of income in relation to the mean across time for each person

Brazil measures monthly income and is quite volatile Ex: Real Minimum Wage in **times of inflation**



Like a Between X Within groups Decomposition

Theil monthly = **Theil 4-month average** + Theil **Temporal dispersion in relation to the individual men income**

Inequality between people (usual)

Variability across Time

Each person is like one group of several observations across Time.

A

		THEIL-T INDEX						
	Population Concept - Income Concept	**1985**	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**
Theil total	**Always Occupied - Month by Month**	0.504	0.651	0.709	0.787	0.533	0.545	0.547
Theil media 4 meses	**Always Occupied - Mean Earnings**	0.448	0.580	0.551	0.646	0.497	0.508	0.512
Theil dispersão de renda média	residuo inst temporal	0.056	0.071	0.158	0.142	0.037	0.037	0.035

Share in Total Inequality: Mean Across People and Across Time around Mean (Same People)

Participação na desigualdade total %

		THEIL-T INDEX						
	Population Concept - Income Concept	**1985**	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**
Theil total	**Always Occupied - Month by Month**	100	100	100	100	100	100	100
Theil media 4 meses	**Always Occupied - Mean Earnings**	88.806	89.069	77.704	82.019	93.086	93.220	93.563
Theil dispersão de renda média		11.194	10.931	22.296	17.981	6.914	6.780	6.437

Dynamic Aspects of Income Distribution withy *Pesquisa Mensal do Emprego* (PME) - This monthly employment survey was carried out in the six main Brazilian metropolitan regions by IBGE. It has covered an average of 40,000 households monthly since 1980. PME replicates the US Current Population Survey (CPS) sampling scheme attempting to collect information on the same dwelling eight times during a period of 16 months. More specifically, PME attempts to collect information on the same dwelling during months t, t+1, t+2, t+3, t+12, t+13, t+14, t+15. This short-run panel characteristic of PME allows us to infer a few dynamic aspects of reforms regarding income distribution.

We have used the micro-longitudinal aspect of PME in two alternative ways: first, the four consecutive observations of the same individuals were treated independently before the inequality measures were assessed; second, we considered earnings average over four months before the inequality measures were calculated. The Theil-T is decomposed as follows: Month by Month Theil-T equals Mean Earnings Theil-T plus Individual Earnings Over Time Theil-T. In other words, the difference in the levels of inequality measures between month by month and average over four months is explained by the variability component of individual earnings over the four-month period.

The main result here is that the fall of month-to-month inequality measures observed after the fall of inflation in 94 drastically overestimates the fall of inequality when one compares it with mean earnings over four months. A comparison of the two lines in TableA indicates that for the always occupied population the month-by-month Theil-T indices fell from 0.709 in 1993 to 0.545 in 1997. The fall of inequality measures based on mean individual earnings over four months is much smaller than in the case of monthly earnings. Theil-T falls from 0.551 to 0.508 between 1993 and 1997. Similar results were obtained for the Gini Index and two other population concepts, such as the active age population and individuals occupied at least once in four consecutive observations, as shown in the paper.

The greater fall of traditional inequality measures on a monthly basis in comparison with measures on a four-month basis is explained by the fall of the individual volatility measures following the sharp decline in inflation rates observed in this period. In sum, stabilization produced more stable earnings trajectories (i.e., lower temporal inequality (in fact, volatility) of individual earnings). On the other hand, the observed fall of inequality *stricto sensu* was much smaller than inequality measures based on monthly measures would have suggested. In sum, the post-stabilization fall in inequality for the group of population always occupied is much higher on a monthly basis (as traditionally used in Brazil) than when one uses mean earnings over four months. The fall of Theils (and Ginis) is 2 to 4 times higher when one uses the former concept.

Another way of looking at the effects of inflation and stabilization is to note that most of the fall in inequality measures is attributed to the within groups component, especially in the month-by-month inequality measures. Table below summarizes this information in terms of the gross and marginal contribution of different groups' characteristics. For example, in the case of the month-by-month income concept presented in part B of table 6.3, during 1993 the sum of the marginal contributions of the between groups component relative to schooling, working class and age (i.e. the three main characteristics) explains only 31.5% of total inequality. This statistic rises to 42.3% in 1997, which corresponds to a 34.3% increase of relative contributive power to total inequality. In the case of the corresponding measures based on mean earnings over four months presented in table 6.3. part A, the relative rise of explanatory power is 12%. These results see to confirm the idea that the explained share of total inequality tends to increase as we approach the permanent income concept. Overall, the main point here is that most of the monthly earnings inequality fall observed after stabilization may be credited to a reduction of earnings volatility and not to a fall in the permanent income inequality (or *strictu sensu* inequality**).**

GROSS AND MARGINAL RATES OF CONTRIBUTION THEIL-T

Universe : Longitudinal Data - 4 Observations - Always Occupied
Mean Earnings Across 4 Months

	GROSS							MARGINAL						
	1985	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**	**1985**	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**
Groups:														
Gender	6.5%	4.4%	3.7%	3.4%	3.6%	3.5%	3.4%							
Age	9.7%	8.7%	7.1%	6.7%	9.1%	9.2%	9.0%	10.4%	7.0%	6.3%	5.7%	6.9%	7.1%	7.6%
Schooling	34.5%	35.8%	32.2%	30.7%	37.5%	38.7%	37.8%	31.5%	30.7%	28.8%	26.8%	32.5%	33.2%	33.1%
Working Class*	10.7%	10.5%	9.2%	11.0%	11.8%	11.8%	12.2%	5.2%	4.5%	5.4%	6.3%	5.7%	5.2%	5.8%
Sector of Activity*	3.4%	2.7%	2.2%	2.3%	1.7%	2.0%	2.1%							
Region	1.6%	2.0%	3.2%	7.0%	4.9%	4.3%	3.3%							

Source: PME
*** Individuals that changed status are classified as Not Specified**

Universe : Longitudinal Data - 4 Observations - Always Occupied
Month by Month Labor Earnings

	GROSS							MARGINAL						
	1985	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**	**1985**	**1990**	**1993**	**1994**	**1996**	**1997**	**1998**
Groups:														
Gender	5.8%	4.0%	2.9%	2.8%	3.4%	3.3%	3.2%							
Age	8.6%	7.8%	5.5%	5.5%	8.4%	8.6%	8.5%	9.3%	6.2%	4.9%	4.7%	6.4%	6.6%	7.1%
Schooling	30.6%	31.9%	25.0%	25.2%	34.9%	36.1%	35.4%	27.9%	27.4%	22.4%	22.0%	30.2%	30.9%	31.0%
Working Class*	9.5%	9.3%	7.2%	9.0%	11.0%	11.0%	11.5%	4.6%	4.0%	4.2%	5.2%	5.3%	4.8%	5.4%
Sector of Activity*	3.0%	2.4%	1.7%	1.9%	1.6%	1.9%	2.0%							
Region	1.4%	1.8%	2.5%	5.8%	4.5%	4.0%	3.1%							

Source: PME
*** Individuals that changed status are classified as Not Specified**

*** A. CONCEPT AND DEDUCTION OF THEIL INDEXES

Reference: *** Hoffmann chapter4 pgs 99 to 116 and c.3 pgs 42-44 (section 3.4). ₴ Theil (1968)

1. Information content of a message

- Based on information theory (Theil (1968) on information content of a message.
- This content depends on the probability of an event occurrence.

Ex: p= 1 => "the event occurred" message has low informative content
p= 0 => "the event occurred" message has high informative content

- Formula

$$h(x) = \log \frac{1}{x} = -\log x$$

- Units

$\log_2 x$ => binary => Bits

$\log_e x$ => natural => Nits (=ln x)

- Examples

Given the rainfall series x = 0,2

$$h(x) = \ln \frac{1}{0{,}2} = 1{,}6094\,Nits$$

Given the rain information in the previous eve y=0,6

$$\mathrm{h}(\mathrm{y}) = \ln \frac{1}{0{,}6} = 0{,}5108\,Nits$$

The information content of the uncertain message (forecast) in question is

$$\mathrm{h}(\mathrm{x}) - \mathrm{h}(\mathrm{y}) = 1{,}0986\,Nits$$

2. Entropy of a distribution

$$\mathrm{H(x)} = \mathrm{E[h(x_i)]} = \sum_i \mathrm{x_i h(x_i)} = \sum_i \mathrm{x_i} \ln \frac{1}{\mathrm{x_i}} = -\sum_i \mathrm{x_i} \ln \mathrm{x_i}$$

We have the following problem:

$$\mathrm{Max\,H(x)}$$

$$\mathrm{s.a.} \sum \mathrm{x_i}$$

$$\mathrm{Max}\{-\sum_i \mathrm{x_i} \ln \mathrm{x_i} - \lambda(\sum_i \mathrm{x_i} - 1)$$

$\mathrm{FOC}: \ln \mathrm{x_i} = -(1+\lambda)$ and the lower bound does not exist but as $\lim \mathrm{x_i} \ln \mathrm{x_i} = 0$ when xi goes to 0

The H(y) maximum, that is, maximum entropy, occurs when there is a maximum of uncertainty about what can happen, once entropy is the expected informative content of a message. This maximum occurs when all possible events are equally probable, and you don't derive any information about those events: $0 \le \mathrm{H(x)} \le \ln n$. The Expected Information of an Uncertain Message is $= \sum_{i=1}^{n} y_i \log y_i / xi$ **which nests the particular full certainty case**

3. Theil Inequality Measures

Theil (1967) proposed an inequality measure from the entropy of a distribution. However, equality do not mean economic disorder (unpredictability). Therefore, he proposed the following transformation: subtracting from entropy its maximum value, we have:

$$T = \log n - H(y) = \left(\sum_{i=1}^{n} y_i\right)\log n + \sum_{i=1}^{n} y_i \log y_i = \sum_{i=1}^{n} y_i\left[\log n + \log y_i\right] = \sum_{i=1}^{n} y_i \log n y_i$$

$$T = \sum_{i=1}^{n} y_i \log n y_i$$

$0 \le T \le \ln n$, that is, we have $T = 0$ in the case of a perfect egalitarian distribution and $T = \ln n$ in the case of maximum inequality.

In the case of $y_i = 0$ we have $y_i \log y_i = 0$, by convention.

where $\mathrm{y_i}$ => share of i in total income

intuitively,

$$\mathrm{T} = \ln n - \mathrm{H(x)} = \sum_i \mathrm{y_i} \ln \frac{\mathrm{y_i}}{1/n}$$

That is, Theil-T index assess how much a given income distribution (each person receive y_i of total income) is away of a perfect uniform distribution (each person receive 1/n of total income), or the redundancy degree in relation to the latter, weighting each observation by its share in total income.

Therefore, the Theil-T index is defined by the following formula:

$$T = \sum_{i=1}^{n} y_i \log n y_i$$

or, alternatively, by

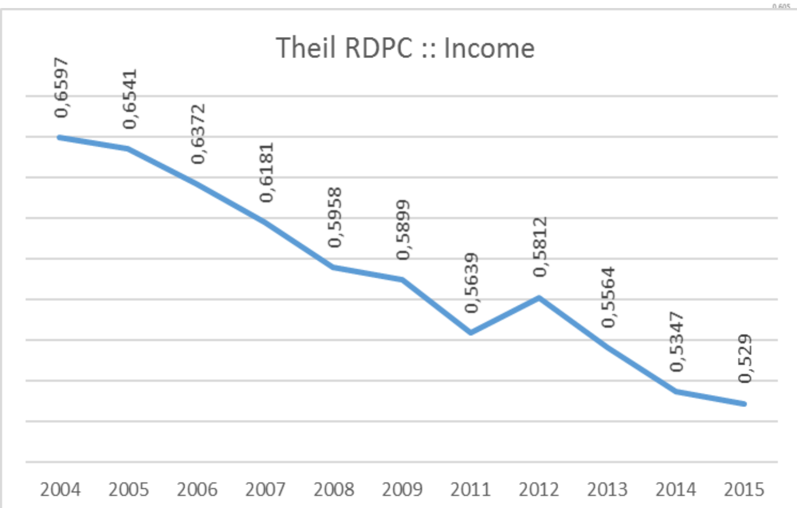
$$T = \sum_{i=1}^{n} \frac{x_i}{N\mu} \log \frac{x_i}{\mu}$$

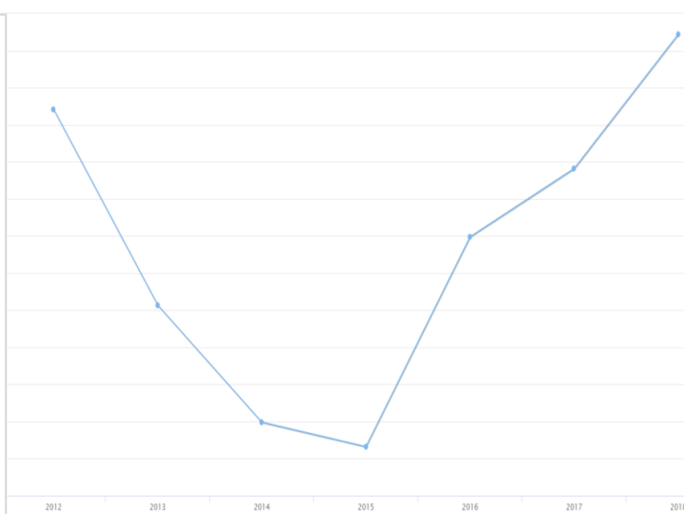
The second Theil measure of inequality is Theil-L index, defined by the following formula:

$$L = \sum_{i=1}^{n} \frac{1}{n} \log \frac{1/n}{y_i} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{1/n}$$

or, alternatively, by

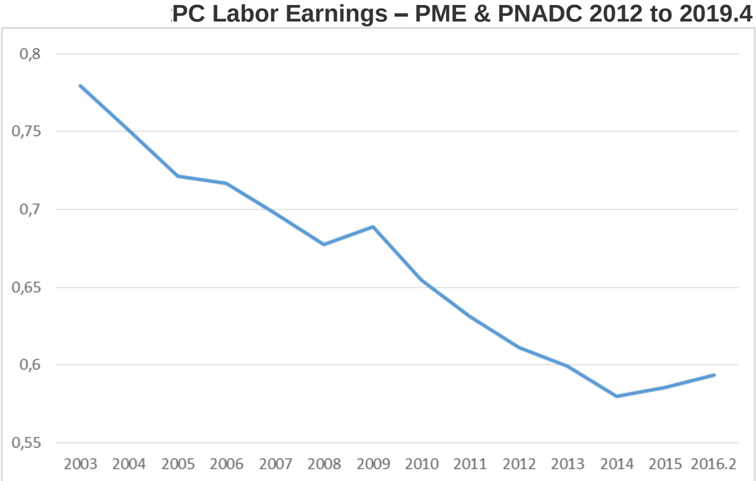
$L = \sum_{i=1}^{n} \frac{1}{N} \log \frac{\mu}{x_i}$ while in Theil T the inequality factors of weighting within the groups are the share of retained income, in Theil L the inequality factors of weighting within the groups are their respective population.

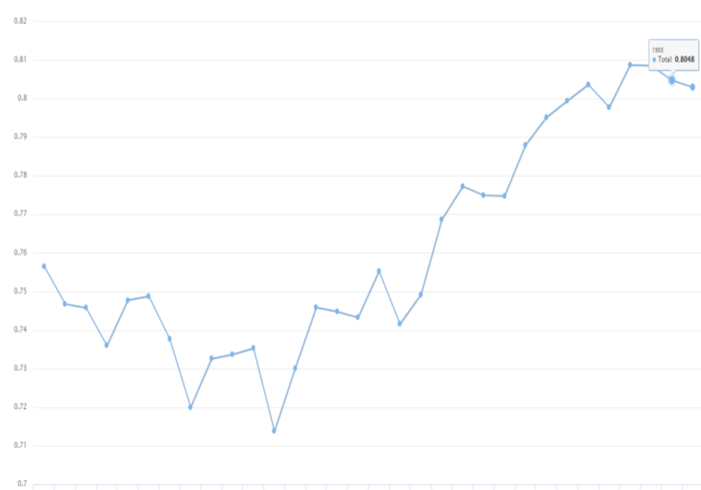




Source: FGV Social based on microdata PNAD 2004-15 and PNADC Annual /IBGE 2012-18 – Per Capita Income All sources

PC Labor Earnings – PME & PNADC 2012 to 2019.4 ->





Source: FGV Social based on microdata from PME /IBGE and PNADC Quarterly /IBGE 2012.Q1 to 2019Q4